NORMAL HUMAN AGING:

The Baltimore Longitudinal Study of Aging

NATHAN W. SHOCK

Richard C. Greulich
Reubin Andres
David Arenberg
Paul T. Costa, Jr.
Edward G. Lakatta
Jordan D. Tobin

U. S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service
National Institutes of Health
National Institute on Aging
Gerontology Research Center
Baltimore City Hospitals

NIH Publication No. 84-2450
November, 1984

CHAPTER III

Design and Operation of the Baltimore Longitudinal Study of Aging

HISTORY

The program on aging of the National Institutes of Health (NIH) was originally established at the Baltimore City Hospitals (BCH) in 1940 (Stieglitz, 1940; Shock, 1947, 1980). Studies of age differences in the performance of selected physiological systems (cardiovascular, pulmonary, renal, muscular, nervous, endocrine, etc.) had used as subjects men aged 60+ who were residents of a domiciliary home for the aged (Infirmary) located on the grounds of the BCH. Since admission to the facility was based primarily on socioeconomic need rather than on health status, the group included residents without apparent disease. Along with staff members, these subjects, after clinical screening to exclude individuals in whom evidence of specific disease could be detected, constituted the population for cross-sectional studies on aging in humans conducted between 1941 and 1958.

At the same time, the desirability of collecting data on non-institutionalized subjects was clearly recognized, as was the fact that a proper study of human aging would require the characterization of age changes in a wide array of physiological, psychological, and social variables across the entire adult age range. Even as late as 1955, most of the literature reporting the effects of age on physiological and psychological variables in man was based on measurements of differences between young adults, usually college students, and older subjects who were frequently residents of hospitals or homes for the aged. Some of the conclusions about human aging derived from such studies were defective or erroneous, since they had compared presumably healthy and active young people with old people who were neither necessarily healthy nor active. Nor did this type of cross-sectional assessment permit an appreciation of the likely influence of birth-cohort effects on the differences observed between the two age groups.

In recognition of the need for a far more systematic approach to the understanding of age changes in man, a plan for a longitudinal study was developed requiring the recruitment of reasonably healthy and active community-dwelling people of all ages, who would be willing to undergo thorough and repeated testing over a major portion, if not the remainder, of their lives. The success of such an enterprise would obviously depend on access to a population of subjects who felt so challenged by their participation that they would voluntarily continue to return for repeated testing.

Our chance meeting with Dr. W.W. Peter provided the necessary access to a group of scientists and educators who had homes or summer cottages at Scientists' Cliffs, a section of the western shore of the Chesapeake Bay located about 60 miles southeast of Baltimore. Dr. Peter, a retired United States Public Health Service officer, was struck by the lack of data on the effects of age in normal people. Moreover, he expressed the strong conviction that people like himself, still living independently, should volunteer to serve as subjects for the study of aging. Thus was born the Baltimore Longitudinal Study of Aging (BLSA).

On his own initiative, Dr. Peter first recruited his friends to serve as volunteer subjects for the BLSA (Norris and Shock, 1960). His circle of friends, however, though large in number, would not provide sufficient subjects to conduct meaningful studies. Hence a plan was instituted whereby each participant actively recruited others among his friends, neighbors, relatives, and colleagues. This plan yielded so many subjects that a waiting list was quickly formed, from which subjects were called as facilities for testing them became available.

During the early stages of the study only two to four subjects could be tested each week, but as facilities expanded more subjects were tested, so that by 1968 about 600 men were actively participating. At this time the aging program was moved into a new building, located on the BCH campus, which was financed by Federal funds and designated as the Gerontology Research Center (GRC) (Shock, 1968).

Since its inception in 1940, the gerontological research program in Baltimore has operated within the administrative framework of NIH, principally as a unit of the National Heart Institute (1948-1963) and of the National Institute of Child Health and Human Development (1964-1975). After the National Institute on Aging (NIA) was established in 1974, all research activities housed at the GRC, including the BLSA, were transferred to NIA, and they have since constituted NIA's intramural (in-house) research program.

OBJECTIVES

From its initiation, the BLSA has had as its goal the systematic description of the processes of aging in humans. In pursuit of that broad goal the study has identified and sought certain objectives:

To describe, both cross-sectionally and longitudinally, the physiological and psychological effects of aging in persons who live in their communities.

To collect serial observations over a period long enough to permit statistically reliable calculation of rates of change in specific variables in individual subjects, and to identify individual patterns of age changes.

To include a broad spectrum of tests applied to the same subjects to determine relations among variables that might lead to an answer to the question, "Is there a general aging process common to a number of physiological and psychological processes, or is aging the end result of multiple independent processes?"

To separate the effects of aging from the effects of physical and mental disease on selected aspects of physiological and behavioral performance.

To assess the influence of age on progressive changes associated with diagnosed diseases.

To examine the effects of critical events, such as loss of job, retirement, death of spouse, or changes in life style or habits, on subsequent tests of physiological and psychological performance.

To develop predictors of age at death and to determine risk factors not only for specific diseases but also for other definable end points, such as loss of mobility, loss of independence, and institutionalization.

To attempt to develop indices of physiological age.

OPERATION OF THE BLSA

To provide a firm basis for interpretation of the results, this chapter outlines the design and operation of the BLSA as it has developed over the years in response to the many advances in the concepts and methods of gerontological research that have occurred since 1958. Many decisions have had to be made on the basis of hunches and best guesses rather than of solid fact, and it has often proven necessary to tailor the study to fit the available resources. In retrospect, while it is obvious that some decisions could have been better, others made out of necessity have turned out to the advantage of the study. A good example is the continuing addition of new subjects to the study population as the study progressed, which has made possible sequential analyses of period and birth-cohort effects (see Chapter I).

1. Administration

Both the scientific and the operational administration of the BLSA have been gratifyingly collegial throughout its years of existence. During the early years, decisions affecting the experimental questions to be addressed were made by the Director, Dr. N.W. Shock, in close conjunction with others of the senior scientific staff of the BLSA. These deliberations, usually undertaken in a regularly scheduled meeting, focused not only on the feasibility and scientific merit of the question to be explored, but also on the justification for addressing it to the BLSA population. Not infrequently, especially in the case of proposed cross-sectional studies, it was determined that utilization of BLSA participants was not essential, and that such investigations could be accomplished as readily by the use of other subjects specifically recruited for the purpose.

Such deliberations emphasize the need for constant overview of the uses to which the BLSA population is put. A concerted effort has been made to avoid the introduction of tests or procedures that are unlikely to lead to longitudinal exploration, or that will unnecessarily add to the already crowded test schedule to which BLSA participants are exposed.

Since 1981, the overall operation of the BLSA has been overseen by a Steering Committee, which includes the NIA Scientific Director as chairman along with five senior investigators whose collective research interests and experience embrace both biomedical and behavioral aspects. This committee serves as final arbiter of all new proposals for BLSA research studies. It is also charged with assessment of the readiness of BLSA data bases for analysis, to decide whether certain tests should be terminated and others introduced.

Operational aspects of the study, including scheduling of visits and tests, as well as communication with the BLSA participants, are the responsibility of a central staff. The subjects in the longitudinal study have always been regarded as participants, never as "guinea pigs." Investigators meet subjects at every opportunity to discuss the goals and operation of the study, and one or more staff members lunch with the subjects each day of their visits. In the early years of the study senior staff members joined the subjects for dinner, and in recent years close association has continued. In addition, each subject receives a regular newsletter that lists publications based on data from the longitudinal subjects and calls attention to important topics being studied, lists subjects who have died, and describes projected studies.

Tests identified by staff physicians as inappropriate or hazardous for certain subjects are excluded from their schedules. Similarly, the selection of subjects for

special tests or the decision to exclude them is the responsibility of the scientific staff. Exclusions that are not based on medical considerations usually reflect conflict among test procedures, stemming either from competing time constraints or from incompatibility among tests; an example is the interference of insulin-infusion studies with tests of pituitary function. These decisions serve as the basis on which the test schedule for each subject is developed.

2. Participants

Development of the study panel. It is often difficult to enlist non-institutionalized volunteers in health-related research. It is even more difficult when participation entails spending two days and nights in a hospital setting, when candidates must furnish their own transportation, and when they are asked to return periodically for additional testing. Certain test procedures may prove burdensome, those designed to assess cognitive functioning can be ego-threatening, and still others call for the expenditure of considerable mental or physical effort. Despite these many and varied demands, the enthusiasm shown by the original members and their efforts at recruiting others

Figure III.1. BLSA volunteers Mr. and Mrs. William E. Wyman of Dover, New Hampshire, during a visit to the GRC.

Table III.1. Recruitment Pattern by Age, Cycles A-K

Cycle	Inclusive Dates	Age (yr)					Total
		17–19	20–39	40–59	60–79	80–96	
A	February 1958–June 1961		51	119	86	4	260
B	July 1961–June 1963		38	81	37	2	158
C	July 1963–June 1965	1	23	68	39	3	134
D	July 1965–June 1967	1	21	61	32	2	117
E	July 1967–June 1969		29	48	36	6	119
F	July 1969–June 1971		57	30	41	3	131
G	July 1971–June 1973	2	29	1	10	3	45
H	July 1973–June 1975		52	2	1	5	60
I	July 1975–June 1977	1	44	14	4	1	64
J	July 1977–June 1979		10	4	7		21
K	July 1979–June 1981	1	15	6	8	3	33
Total		6	369	434	301	32	1142

contributed materially to the successful initiation and continuation of the BLSA. Figure 1 depicts two of the volunteers who have made the study possible.

Over the years, a few subjects joined the program at the invitation of staff members, and some volunteered after reading or hearing about the study. More than 90% of participants, however, were proposed for membership by friends or relatives who were already taking part in the study. In consequence, most new members have at least some familiarity with the nature of the testing program by the time of their first visit, and because many participants elect to return to the GRC together their visits tend to be pleasant social occasions.

During the initial visit, the nature and goals of the study are explained to each subject, and special emphasis is placed on the fact that, while the program is not designed to provide medical services, the results of the clinical and laboratory tests will be referred to each participant's personal physician. When the study started, a letter describing the tests to be performed was sent to the physician, who was asked to report directly to the GRC if, in his opinion, any of the tests proposed would present a hazard to his patient. Since none of the physicians raised objections to any of the tests, the letters were discontinued in 1962.

All subjects sign a statement of informed consent, which specifies the nature of all procedures and tests to be used, indicates potential hazards, and clearly states that each subject has the right to decline participation in any of the tests. New test procedures are reviewed and approved by the BCH Institutional Review Board.

In order to facilitate data analysis the repeated observations were grouped in 24-month periods, or cycles, designated alphabetically. Table 1, which shows the recruitment pattern by age groups for cycles A through K, reveals that recruitment of new subjects was most active between 1958 and June 30, 1971. Table 2 reflects the age composition of the panel at the end of each cycle.

Table 3 shows the total number of subjects who had been tested by June 30, 1981, distributed by age at time of first visit and number of visits. By that date 1142 subjects

Table III.2. Panel Age Composition, Cycles A-K

Cycle	Inclusive Dates	Age (yr)					Totals	
		17–19	20–39	40–59	60–79	80–96	Subjects	Visits[a]
A	February 1958–June 1961		51	119	86	4	260	369
B	July 1961–June 1963		69	175	119	7	370	445
C	July 1963–June 1965	1	73	214	143	12	443	536
D	July 1965–June 1967	1	64	272	178	20	535	646
E	July 1967–June 1969		62	287	201	22	573	793
F	July 1969–June 1971		106	292	239	31	668	913
G	July 1971–June 1973	2	99	262	226	34	623	784
H	July 1973–June 1975		135	227	223	43	628	776
I	July 1975–June 1977	1	156	217	218	39	631	750
J	July 1977–June 1979		121	205	228	41	595	691
K	July 1979–June 1981	1	118	196	222	48	585	695

[a] The number of visits within a cycle exceeds the number of subjects because BLSA participants over 60 years of age visited GRC more than once during the 2-yr cycles.

Table III.3. Cumulative Number of Male Subject Visits by Age at First Visit[a] (as of June 30, 1981)

Number of Visits	Age at First Visit (Yr)									Total
	17–19	20–29	30–39	40–49	50–59	60–69	70–79	80–89	90–96	
1	6	143	226	242	192	153	148	27	5	1142
2	4	128	202	217	175	136	121	20	3	1006
3	4	101	176	201	155	123	104	16	3	883
4	3	77	145	179	145	113	90	11	3	766
5	1	49	123	173	135	100	76	7	3	667
6	1	31	109	161	122	96	60	6	1	587
7	1	21	88	154	107	88	55	6		520
8		15	74	133	91	79	52	5		449
9		12	64	113	79	68	44	4		384
10		9	48	87	62	58	33	4		301
11		9	32	68	51	52	26	2		240
12		3	17	45	39	39	19	1		163
13		1	8	24	28	29	15			105
14		1	2	12	20	26	9			70
15			2	6	13	19	5			45
16			1	2	10	14	3			30
17			1	2	6	9	2			20
18			1	1	2	5	2			11
19			1		1	3	1			6
20						2				2
21						1				1

[a]As a guide to interpretation of this table, it may be noted, for example, that 122 of the 192 subjects who began their participation in the BLSA when they were between 50 and 59 yr of age have made 6 visits or more since their entry into the study. Similarly, 2 of the 27 subjects who were between 80 and 89 at entry have made 11 visits, and one of these 2 had a 12th visit. The last column indicates, for example, that 240 subjects of all ages have made 11 visits.

had been seen one or more times, and nearly half of these, 520, had made seven or more visits.

While no systematic effort has been made to identify the motives that have prompted subjects to volunteer, it is clear from casual conversation that the opportunity to obtain a comprehensive physical examination at regular intervals has been an important factor. Nevertheless, since virtually all participants have personal physicians to whom medical reports are sent, and most are able to afford medical care, the physical examination cannot be more than partial compensation for the many demands of participation. It is obvious that a desire to contribute to aging research has been a large part of the motivation.

Although in the early years of the study subjects were not selected on the basis of age, after July 1970 preference was given to subjects in certain age decades in order to maintain at least 50 subjects in each five-year age group from 25 to 85 (Tab. 1). Most of the subjects live in the Baltimore-Washington metropolitan area, although some, especially those who have retired, return from such states as Michigan, Florida, New Hampshire, Maine, Arizona, and California.

When the study was initiated the resources available to the GRC for overnight housing of participants consisted of a single room, so that bathroom facilities had to be shared with ambulatory elderly male patients in a large ward of the BCH. It was thus necessary to limit the longitudinal study to men. With the opening of the GRC building and renovation of the ward space in the hospital in 1968, physical facilities became available for housing women. Because of limitations of staff and budget, however, recruitment and examination of women did not begin until 1978.

Socioeconomic and educational characteristics and health status of the participants. The socioeconomic characteristics of the sample generated by this method of recruitment are those of an upper-middle-class segment of the general population. An assessment of the social characteristics of subjects recruited up to June 30, 1981, is presented in Table 4. As the table reveals, 84% of the subjects were identified at initial visit with professional, technical, and managerial occupations, 71% had bachelors' or higher academic degrees, and 73% (82% of those reporting) rated their financial situations as comfortable or better.

In addition, the sample includes a high proportion of government employees: 56% were or had been in municipal, state, or federal employment, exclusive of public-school teaching and military service.

The sample at entry (Tab. 5) was also relatively homogeneous with respect to reported health, race, marital status, and religion. That 85% (93% of those reporting) rated their health as good or excellent is unremarkable, since one of the objectives of the study was the investigation of healthy aging. No subject was denied participation for reasons of health; nor were criteria of race or social status applied to restrict admission to the program.

It is conceivable that men of different ages or generations might have volunteered for different reasons, with the result that the sample would have varied in its social composition by age. This source of bias seems largely to have been avoided, however, in that the socioeconomic attributes listed in Table 4, and the distributions of health, race, and marital status by age in Table 5, either reveal a substantial correspondence among age groups or reflect changes in financial, health, and marital status that are usually associated with older age groups.

Although above-average educational level and socioeconomic status were not

Table III.4. Percentage Distribution of Male Subjects by Age and Socioeconomic Characteristics at First Visit (as of June 30, 1981)

	Age at First Visit (Yr)					
	17–19	20–39	40–59	60–79	80–96	Aggregate (17–96)
Number of Subjects	6	369	434	301	32	1142
Occupational Class						
Professional		56.9	66.6	67.1	71.9	63.4
Managerial		18.4	23.0	18.3	21.9	20.1
White collar		10.0	4.6	7.6	3.1	7.1
Blue collar		13.8	5.8	7.0	3.1	8.6
Student	100.0	0.8				0.8
	100.0	99.9	100.0	100.0	100.0	100.0
Educational Status						
Less than baccalaurate degree	100.0	36.3	23.7	25.2	31.2	28.8
Baccalaurate degree		32.2	27.6	24.9	18.8	28.0
Master's degree		20.6	22.4	14.6	21.9	19.6
Doctoral degree		10.8	26.3	35.2	28.1	23.6
	100.0	99.9	100.0	99.9	100.0	100.0
Present Economic Status						
Can't make ends meet		1.9	0.5	0.3		0.9
Enough to get along		22.0	13.8	10.6	9.4	15.4
Comfortable	50.0	55.0	67.7	68.4	53.1	63.3
Well-to-do	16.7	5.4	9.0	16.3	9.4	9.8
No information	33.3	15.7	9.0	4.3	28.1	10.6
	100.0	100.0	100.0	99.9	100.0	100.0

specifically applied as criteria for admission to the study, the method of recruitment resulted in a population possessing both attributes. The participation of highly educated subjects offered a number of advantages for the study of aging: improved accuracy of such information as family history, past history of illnesses, current medical status, and medication being taken; accuracy of activity history; accuracy of nutritional diaries and ability to follow detailed instructions in recording dietary intakes; ability to understand and execute a variety of questionnaires; and increased probability that the subject would be interested in the scientific questions addressed by the study. In short, this selection of subjects provided the opportunity to study the effects of aging under conditions in which economic status and educational level would not seriously limit medical care, nutrition, and other factors affecting health over the life span. The results of the study may thus be considered to reflect the effects of aging under optimal socioeconomic conditions.

Stability of the sample and losses from the program; drop-outs. In spite of additions to and losses from the study population over the history of the program, the social characteristics of the sample have remained surprisingly stable. Table 6 compares groups tested during each two-year cycle with respect to various attributes at entry into

Table III.5. Percentage Distribution of Male Subjects by Age and Personal Attributes at First Visit (as of June 30, 1981)

	Age at First Visit (Yr)					
	17–19	20–39	40–59	60–79	80–96	Aggregate (17–96)
Number of Subjects	6	369	434	301	32	1142
Health Status						
Poor		0.5	1.2	0.7	3.1	0.9
Fair		3.0	5.5	9.6	6.2	5.8
Good		33.3	47.2	50.5	40.6	43.2
Excellent	66.7	50.1	39.4	35.5	21.9	41.5
No information	33.3	13.0	6.7	3.6	28.1	8.7
	100.0	99.9	100.0	99.9	99.9	100.1
Marital Status						
Never married	100.0	17.1	2.3	1.7	3.1	7.4
Married		77.8	94.9	89.0	53.1	86.2
Formerly married		5.1	2.8	9.3	43.8	6.4
	100.0	100.0	100.0	100.0	100.0	100.0
Religious Affiliation						
Roman Catholic	33.3	22.8	14.7	8.0	3.1	15.3
Jewish		7.0	7.6	3.0	3.1	6.0
Protestant	50.0	59.1	72.6	81.7	84.4	70.8
Other	16.7	8.1	4.1	6.6	9.4	6.3
No information		3.0	0.9	0.7		1.5
	100.0	100.0	99.9	100.0	100.0	99.9
Color						
White	100.0	95.1	96.6	98.3	96.9	96.6
Black		4.9	3.2	1.7		3.2
Other			0.2		3.1	0.2
	100.0	100.0	100.0	100.0	100.0	100.0

the program. Minor changes have occurred during the 23 years. Mean age rose from 53.0 years in Cycle A (1958–1961) to 57.0 years in Cycle K (1979–1981). Over the same period the proportion of subjects identified with professional, technical, and managerial occupations fell from 91% to 84%, and Protestant affiliation from 79% to 71%. Sample composition remained remarkably uniform over all cycles in the proportions of participants with bachelors' or higher degrees at first visit.

Although most subjects who joined the study were aware of its demands, some found it difficult to continue. While no final assessment of the causes of their withdrawal has yet been made, predictors include the status of the recruiter (whether the person who recruited the subject is still in the program), personality traits reflecting critical dispositions—both specific complaints about tests or staff members and more

Table III.6. Percentage Distribution of Male Subjects by Mean Age and Selected Attributes at Each Cycle[a] (Total N in Parentheses)

Characteristic	Cycle										
	A	B	C	D	E	F	G	H	I	J	K
Occupational Class											
Professional, technical, or managerial	91.2 (260)	90.3 (370)	89.2 (443)	88.0 (535)	86.7 (573)	85.0 (668)	84.8 (620)	84.6 (604)	84.1 (603)	83.9 (571)	84.1 (552)
Education											
Bachelor's or higher degree	79.6 (260)	78.9 (370)	79.7 (443)	78.3 (535)	77.5 (573)	77.7 (668)	77.7 (623)	75.3 (628)	74.0 (631)	74.0 (595)	72.7 (585)
Health (Self-rating)											
Good or Excellent	93.7 (253)	94.1 (358)	93.0 (428)	92.6 (516)	93.4 (545)	93.0 (632)	93.8 (592)	94.1 (580)	95.8 (575)	95.1 (551)	95.7 (532)
Religious Affiliation											
Protestant	78.9 (256)	78.2 (327)	77.5 (440)	76.6 (533)	78.5 (572)	74.5 (667)	75.3 (619)	73.4 (601)	72.1 (599)	71.0 (567)	70.7 (549)
Mean Age	53.0 (260)	53.8 (370)	54.6 (443)	55.7 (535)	56.3 (573)	55.8 (668)	56.6 (623)	55.6 (628)	54.9 (631)	56.7 (595)	57.0 (585)

[a] Attributes other than age determined at first visit only.

diffuse ones about the program and its environment—and changes in personal circumstances such as illness of the subject or of members of his family.

When subjects who had not returned after one or two visits by June 30, 1977, were compared with respect to age, occupation, highest degree, religious affiliation, and financial status at first visit with those who had failed to return after three or more visits, no significant differences between the two groups were observed. The drop-outs as a group (N = 280) were also found to be similar, except in the level of academic degrees, to subjects who had made three or more visits and were still active as of June 30, 1977 (N = 658). Fewer drop-outs than active subjects had held masters' degrees and doctorates at first visit.

As of June 30, 1977, 1088 subjects had been tested at least once. At that time a major follow-up study was made of the participants who had formally withdrawn or had not returned in a three-year period. Table 7 summarizes the results. By the cut-off date, 150 subjects had died and 280 had failed to return within three years. Of the 280 all but five were located. Of the 275 subjects located, 56 had died since their last visit. A priority cascade was established for the 219 who were still alive. Our top priority was to persuade them to return for additional visits. Since several of the major reasons for dropping out—serious illness, institutionalization, distant move—made return difficult or impossible, we were pleased that 76 (35%) returned. As the next step in our follow-up, we asked that a home visit be permitted, and were rewarded with a further 19%.

Table III.7. Follow-up Study of Male Subjects Seen at Least Once from February 1958 to June 1977

Total subjects seen	1088		
Active as of June 1977	658		
Died while an active participant	150		
Failed to return within 3 years of last visit ("Drop-out")	280 →	280	
Could not be located		5	
Located by follow-up		275 →	275
Died after leaving study			56
Alive at time of follow-up			219 → 219
Re-enrolled in study			76
Visited at home			42
Responded to mailed questionnaire			71
Responded to telephone interview			9
Subject undecided			19
Refused further cooperation			2

Eighty subjects (37%) completed a written questionnaire or responded to a telephone interview with respect to their health status. There is still the possibility of some follow-up on 19 subjects (9%) who have not yet decided how to respond. Only two have refused further cooperation. We thus know whether the subject is still living in 99.5% of the cases, and health and social information is available in 97.6%.

3. The Tests

This section relates the administration of tests to the purposes and design of the study. A detailed description of the tests administered is presented in Chapter IV.

Criteria for selection of tests. The battery of tests used in the study was designed to cover a broad spectrum of physiological, psychological, and social variables in order to permit the evaluation of relations among them and their association with aging. In addition to tests of the performance of specific organ systems, such as the heart, lungs, kidneys, and muscles, tests of the ability to integrate the activities of a number of organ systems in adapting to physiological stresses, such as exercise, were included. Other tests, such as glucose tolerance, measured the rates of displacement and recovery of physiological variables.

Selection of tests for repeated administration was based on the following considerations: a) the probable importance of the variable in understanding age changes in performance and adaptability; b) evidence from cross-sectional studies that age differences are present in the function; c) established validity and reliability of the test; d) the presence on the staff of an investigator interested in the test and qualified to administer it and interpret the results; and e) the presence of minimal risk and unpleasantness to the subject.

Testing schedule. At each visit subjects spend 2½ days at the GRC. Meals are provided in the cafeteria of the BCH, but subjects frequently elect to take one of the evening meals outside the hospital. The length of visit was chosen to provide sufficient

time for a diversity of studies, including tests such as basal metabolism and glucose tolerance that must be made under basal conditions. The visit schedule also permits the collection of 24-hour urine samples required for the determination of creatinine clearance, the excretion of hormones, etc., as well as the introduction of lengthy and complicated laboratory procedures.

During each visit subjects rotate through various laboratories on a schedule that is prepared in advance for each subject. Individual testing schedules are also based on the time that has elapsed since the performance of other tests that might produce interference with the results. Tests administered during the course of the study fall into four general categories:

1) Tests administered to all subjects on all visits. Examples include medical history and physical examination, anthropometry, basal blood pressure and ECG, vital capacity, maximum breathing capacity, creatinine clearance, and blood chemistry (fasting cholesterol, triglycerides, and fasting glucose).

2) Tests administered regularly to all eligible subjects, but not at each visit. Some tests cannot be administered together in the same visit because of interference caused by the administration of drugs or hormones, the amount of blood withdrawn, or the time required. The frequency of administration of particular tests was determined on the grounds of availability of investigators and subject time, interference with other tests, and length of the period over which age change might be expected to occur.

3) Tests administered at only one visit, primarily to assess stable characteristics of the subjects, include determinations of such genetic markers as blood groups and taste sensitivity to phenylcarbamate.

4) Tests introduced for specific approved investigations that, although not necessarily longitudinal in design, could be carried out with maximum efficiency and minimal cost because a great deal of background information about the subjects from the longitudinal study was available. The study of age differences in the physiological and psychological effects of ethanol is an example.

During the early years of the study subjects traveled to Baltimore in pairs. At first only two pairs of subjects could be tested each week, but as resources expanded the number grew with them, so that by June 1968, 12 subjects were scheduled each week: six to arrive at the GRC each Monday and six each Wednesday morning. Subjects arriving Monday morning left at noon on Wednesday, and those arriving Wednesday morning left at noon on Friday. In 1978 the schedule was altered so that 15 subjects were scheduled each week: Six arrived for testing on Monday, three on Tuesday, and six on Wednesday mornings. Each group remained at the Center for two days. (This schedule was followed in order to provide time for testing women in the program that began in January 1978.)

In the original design, subjects aged 20 to 70 years were tested at 18-month intervals, while subjects over the age of 70 were seen annually. In July 1970 the test interval was set at 24 months for subjects aged 20 to 59, 18 months for subjects aged 60 to 70, and one year for all subjects over 70 years.[2] Table 2 shows both the number of subjects tested in each cycle and the total number of visits. When subjects were tested twice in a cycle, only the data from the first visit were used for cross-sectional analysis, although the data from both visits were used in longitudinal analysis.

[2]Since July 1981, all subjects have been seen each two years.

Quality control of test administration. Responsibility for training and supervision of technical assistants, as well as overall quality control of the data collection, rests with each investigator, who also maintains files of original data and reviews summary data before they are introduced into the central data file.

Changes in methods. Improved methods for tests already in use are introduced when they become available. Whenever methods are changed, measurements are made at the same visit by both the old and the new techniques in a series of subjects distributed over the total age span, so that if systematic differences are found in the values obtained by the two methods, appropriate adjustments can be made in the earlier measurements to maintain comparability throughout the longitudinal analysis.

Introduction and elimination of tests. New and expanding interests of investigators, as well as new developments in the field, have led to the incorporation of new tests as the study has progressed. New tests are introduced only after review by the Steering Committee of a written proposal submitted by the investigators. The testing cycle and the age span of the sample make it possible to assess the potential value of a new test by cross-sectional analysis of data collected over a two-year interval.

Over the years of the study, some tests have been eliminated because the study had been completed, because the principal investigator had left the GRC, or because it seemed unlikely that further testing would yield fruitful results.

APPROACHES TO INTERPRETATION OF DATA

1. The Aging-Disease Dilemma

There is common agreement that, in order to characterize true biological age changes in specific physiological systems, measurements in subjects with known diseases or pathology of the organ system under study should be excluded from analysis. Hence the necessity that normal aging be studied in "healthy" or "disease-free" subjects. This requirement has not so far been satisfied. In the first place, for certain conditions it is difficult to certify that a given characteristic is absent; the scientific method can only show that with current methodology a given condition cannot be detected. Since disease may be present but undetectable, it must be assumed that new and more sensitive methods could show its presence.

Unlike some scientists who believe that there is no unique biological process of aging distinct from pathological or disease-related processes, BLSA scientists have adopted the working hypothesis that aging is distinct from disease. Under this assumption great care is taken in the refinement of raw data and the exclusion of data obtained from subjects with specific diseases. An example of this process may be found in our study of creatinine clearance *(Rowe et al., 1976b);* the numerical impact is shown in Table 8. Consider the following example: A major goal of studies of blood pressure currently under way is to describe changes that occur with aging, but to eliminate changes induced by drugs or diseases that would, in themselves, change blood pressure. Variables known or believed to influence either cardiac output or peripheral resistance were considered. Thus, if a subject reported that he was taking any of the following medications, measures from that visit were excluded: diuretics and antihypertensives, amphetamines, methylphenidate (Ritalin^{R}), digitalis, long-acting coronary, cerebral, peripheral, or "general" vasodilators, anti-arrhythmics, oral and injectable adrenal or ovarian steroids, certain non-steroid anti-inflammatory agents, major

Table III.8. Number of Subjects Excluded from Normal Group by Disease Category and Age Group[a]

	Age (Yr)			
	17–24	25–44	45–64	65–96
Nephrolithiasis		8	47	21
Urinary tract infection		17	32	25
Gout		4	10	4
Prostatectomy		2	11	60
Congestive heart failure			2	3
Coronary heart disease		7	45	44
Cerebrovascular disease		2	16	18
Diabetes mellitus		10	36	18
Abnormal urinalysis	1	10	11	16
Miscellaneous renal disease	1	15	17	18

[a] Since some subjects had more than one disease, the total number of individual exclusions (531) in this table is necessarily larger than the number of subjects excluded (336).
From *Rowe et al. (1976b)*

tranquilizers, ergot alkaloids, narcotic analgesics, and L-dopa. Lesser drugs, e.g., minor tranquilizers, antihistamines, and sedatives, have not been shown to affect blood pressure.

The presence of any of the following conditions or diseases leads to exclusion from studies on the relation between age and blood pressure:

- Coronary artery disease. The point system used is described in Chapter IV. Measures are included in age analyses until the diagnosis of coronary artery disease is established, whereupon all subsequent measures are excluded.
- Aortic valvular stenosis or incompetence.
- Definite evidence of cerebral infarction.
- Hypo- or hyper-thyroidism at the time of the visit.
- Diabetes mellitus. Subjects who have ever taken either insulin or oral hypoglycemic agents are excluded. Also excluded are subjects with fasting plasma-glucose values of 140 or more on two or more occasions.
- Cancer with evidence of systemic effects, such as weight loss or anemia.
- Renal diseases, such as polycystic or horseshoe kidney, nephrectomy, or laboratory evidence of uremia (see below), regardless of etiology.
- Miscellaneous diseases, including Fabry's Disease, panhypopituitarism, and Buerger's Disease. Our study included no known cases of pheochromocytoma, adrenocortical disease, or coarctation of the aorta.

There are also laboratory criteria for the exclusion of blood-pressure measurements from estimates of the effects of aging:

- Abnormal urinalysis.
- Twenty-four-hour creatinine clearance with age-adjusted centile ranking of less than 1%, confirmed by age-adjusted serum-creatinine concentration ranking of less than 1%.

- Hemoglobin of less than 11 g/dl, or hematocrit of less than 33% or more than 55%.

One further exclusion is made, not on the basis of drugs, diseases, or laboratory findings, but of the well-known stress effect of the initial blood-pressure reading, the "first-visit artifact" (which in some studies has extended to the second visit as well). An analysis for this variable in our study showed the stress-effect to be evident only on the first visit. Measurements made on the first visit are excluded from calculations of the regressions of blood pressure on age.

2. The Concept of Clinical "Clean-Up"

The process of identifying and excluding subjects with conditions or diseases that might influence the values of a variable under study, or of excluding certain data points, is referred to as clinical clean-up. The necessity of a clinical clean-up preliminary to the analysis of data on aging processes is controversial. Consider the following example:

The creatinine clearance of a subject has been measured annually as he aged from 60 to 70 years. The slope of the 24-hour creatinine clearance on age, computed at age 70, is a measure of the rate of aging of his renal function, or glomerular filtration rate. After his last visit, certain symptoms develop; a nephrectomy is performed to remove a renal carcinoma. When he returns at age 71 a marked drop is of course found in the measured creatinine clearance. If the goal of the study is to define the effects of normal aging processes on selected physiological functions, should this last datum be included? Should a new slope be computed on data collected from age 60 to age 71, or should the assessment of the aging of renal function in the subject be stopped at age 70? In this example, it would appear absurd to disregard the surgical removal of a kidney—or to classify the event as normal aging. Conclusions based on computed slopes of decline that do not take this kind of event into account are bound to be misleading.

But consider a second example. Suppose that investigators studying renal function believe that hypertension may accelerate decline, and that subjects with this disease should be excluded. Since blood pressure increases with age, if a standard cut-off such as systolic pressure above 140 is adopted, the number of very old subjects eligible for inclusion in the study will be severely limited. Moreover, those who remain are likely to represent a "biological élite" whose health status is extraordinary. It is difficult to argue that normal aging can be studied in such a supernormal population. Utilizing a higher, "age-adjusted" cut-off point for older subjects is also unsatisfactory: If systolic pressures above 140 adversely affect the kidneys of 40-year-olds, surely they must adversely affect the kidneys of 80-year-olds. In this example, clinical clean-up may be inadvisable.

There are a number of alternative decisions the investigator may make, depending on his research goals. He may, for example, decide to screen for some medical conditions (such as surgical removal of a kidney), but not for others (such as hypertension), on the presumption that the former are both more directly relevant to the organ system under investigation and less likely to represent normal aging changes.

If the goal is the detection of pure aging effects, then a rigorous screening, leaving only a select group, may be warranted. Results from these studies are particularly useful in establishing that declines in functioning are an intrinsic part of the aging process.

Failure to find age changes would be difficult to interpret, however, since a true decline may be missed because of the bias introduced by the élite sample.

If an investigator is concerned primarily with the generalizability of his findings—if he wishes to describe changes that typically accompany age in the general population—then the best decision may be not to screen at all. This is also the strategy to be preferred if the researcher believes that the variable of interest does not change with age. If he can demonstrate stability despite both aging changes and the illnesses that accompany age, he has made a stronger case than could be made from a screened population.

At a minimum, it is our conviction that the decisions taken by each investigator must be presented in sufficient detail so that results will be properly interpreted and that comparability with other studies may be maintained. There is little question that other investigators faced with the same imperatives of categorization would arrive at a set of rules at some variance from ours. Indeed, candor requires us to admit that our own rules have tended to evolve not in response to an analysis of the end results of different strategies but according to standards of reasonableness (or unreasonableness) that have become evident as the initial set of rules has been applied. To illustrate: In an analysis of blood-pressure changes with age, a preliminary decision was made to consider the presence of a single cast of any type in the urine as indicative of a state of "renal disease" sufficient to exclude data from that subject's visit as "nonhealthy." This turned out to be nonsensical; urinary microscopic reports are replete with findings of single casts, with no ancillary evidence at all of renal disease at that visit or at prior or subsequent ones. When this criterion was modified to two or more casts, the percentage of excludable results dropped precipitously.

Perhaps the best strategy for dealing with this dilemma is to conduct and report analyses both with and without clinical clean-up. A comparison of the results of the two approaches will provide a better basis for both interpretation and generalization. If the same results are found, we can infer that the effect should be considered a manifestation of "aging," and also that it is found in the population in general. If results differ, we must be much more cautious in interpreting them, because either disease processes or sample biases could account for the effects. At this point it might be useful to conduct analyses on subgroups distinguished by particular diseases, to see if a single category of illness can explain the discrepancy. For example, screening for surgical removal of the kidney might make a considerable difference, whereas screening for hypertension might not. In these analyses we could learn not only about aging, but also about the consequences of specific disease states. This information, in turn, might be useful to future investigators who want to know which conditions are most important in a clinical clean-up.

THE DATA-MANAGEMENT SYSTEM

The data-management system of the BLSA provides a flexible facility for collection, storage, and analysis of those physiological, psychological, biochemical, medical, sociological, nutritional, and body-composition measurements described in Chapter IV. The system accommodates a variety of input modes and provides a wide-ranging, continually changing set of summaries and analyses responsive both to the needs of scientists and managers and to the legal requirement that confidentiality be

assured. Scientists require access to a variety of statistical packages in addition to the implementation of their personal methodologies. Manager summaries, reports, projections, and schedules in most cases require programming *de novo.*

Facilities include a Digital Equipment Corporation (DEC) PDP 11/70 computer, which utilizes MUMPS (Massachusetts General Hospital Utility Multi-Programming System) as an operating and programming system and serves as a secure repository of the BLSA data. Scientists and managers gain access to this system through remote terminals in the laboratories and offices for entry, review, editing, and retrieval of single-record data.

The data base comprises 670,000 records, of which 380,000 are in fixed-field format (data in identifiable locations that remain the same for all persons and times). Fixed-field records include scientific results and numerically encoded data from medical histories and physical examinations. Medical diagnoses are recorded on another 60,000 records. The remaining 230,000 records contain verbal responses to specific questions, as well as free comments on fixed-field information. All free-form data are retrievable through numeric subject-matter codes that precede each entry in these fields, and by key-word search.

Batch-processing support for computation and analysis of the data includes a local Digital Equipment Corporation (DEC) VAX 11/780 computer and NIH's Division of Computer Research and Technology in Bethesda, Maryland, some forty miles distant, access to which is provided by a remote job-entry system (Data General Eclipse C-150) at the GRC.

Software is available to extract subsets of the data from the data-base system in forms suitable for analysis by standard packaged statistical programs (SAS, BMDP, SPSS, etc.) or by programs developed specifically for the BLSA by GRC scientists. Examples of the latter include several scanning routines that may be applied to analysis of data from subjects who have made several visits. Trends with time in a given variable, significant deviations from the trend or mean performance level, and regressions on time may all be evaluated by these routines and are essential both to quality control and to preliminary longitudinal analysis.

When data are available for only one or two visits, another program permits comparison of the subject's performance level with an appropriate distribution of similar values. Each subject must be compared with others measured at the same time, of similar age, of a similar disease classification, of similar body composition, height, or weight; or with subjects from another study in the literature in which the distribution of values is sufficiently well described.

Laboratory variables may be plotted against age, height, smoking habit, disease class, or any other variable in the file. This provides a quick and easy comparison of variance, for example, between old and young, tall and short, smokers and nonsmokers, etc. As a tool for cross-sectional analysis, these plots permit a judgment about the appropriateness of a linear or other fit to the data. Longitudinal data can be evaluated by comparison of these relations from one time of the study to another. A variation among the relations shown in the plots of a variable for different times suggests the need for a review of the laboratory results to determine whether longitudinal change, secular trends, or measurement error is responsible for variation over time.

Age regressions are based on a selected number of serial observations for a variable, such as the first five, all available, etc. Individual regressions are summarized in one or more age groups for the purpose of presenting the changes in subjects 35–44,

40–49, 40–59 years of age, etc. Plot programs are available for displaying age-group slopes around mean x, y, with a line length equal to the average period of observation on the time axis.

More specific and sophisticated computer routines for special data analyses, e.g., multivariate analysis, are developed by GRC investigators as the need arises.